A Review of Landsat-7 and EO-1 Cloud Detection Algorithms

By Nargess Memarsadeghi, NASA/GSFC, Code 587

1. Introduction

Cloud detection algorithms play a critical role in many remote sensing studies. Scientists are interested in identifying clouds in satellite imagery for various reasons. First, results of many climate and weather models and analysis heavily rely on the amount and types of clouds that were present in the regions of study. Second, even if clouds are not the objective of a study themselves, they serve as obstacles and sources of errors when sensing other natural phenomena. Thus, scientists often times apply cloud detection algorithms to satellite imagery. The output of such an algorithm is often times called a cloud *mask*. A mask is an image of the same size as the input images of interest, where each pixel can have only two (or a few) values. For example, the simplest cloud mask is a two dimensional image where each pixel has either value 1 if it is classified as a cloud pixel, or 0 otherwise. In a slightly more complicated cloud mask, each pixel can have one of, for example, four different possible values for clear, low, and high clouds each and one for an unclassified pixel.

Different cloud detection algorithms have been developed for various instruments of the Earth science missions. However, they are all developed using the same fundamental reasoning and logic. In this document, we discuss cloud detection algorithms for Landsat-7 and EO-1 Hyperion datasets. These algorithms have some common preprocessing steps, which are described in Section 1.1. We then describe the cloud detection algorithms for Landsat and EO-1 in Sections 2 and 3 respectively. References for this document are listed in Section 4.

1.1 Common Steps in Landsat and EO-1 Cloud Detection Algorithms

Cloud detection algorithms for multispectral data of Landsat-7 and EO-1 Hyperion data have several common steps. For both algorithms, only a few bands corresponding to wavelength of interest are selected and analyzed with the cloud detection algorithm. For example, only 5 bands of Landsat-7 data and 6 bands of Hyperion data are used for cloud detection.

Cloud detection algorithms are sensitive to reflectance values of each pixel in the image. Thus, it is important that the instruments irradiance values be first converted to reflectance values. This step removes the effects of sun angles on the measurements. The reflectance value for pixel number *i* in an image ($\boldsymbol{\rho_i}$) is a function of its measured irradiance value ($\boldsymbol{L_i}$), solar zenith angle ($\boldsymbol{\theta}$), the Earth-Sun distance ($\boldsymbol{d}$), and the mean solar irradiance value ($\mathrm{E}_{\mathrm{SUN}_\lambda}$).

$$\boldsymbol{\rho_i} = \left(\frac{\boldsymbol{\pi d^2}}{\mathbf{cos}(\boldsymbol{\theta})\mathrm{E}_{\mathrm{SUN}_\lambda}}\right)\boldsymbol{L_i} \qquad (1)$$

The solar zenith angle $\boldsymbol{\theta}$, can be obtained from the date and time of the data acquisition, and knowing the location of the observed data on earth. With the required information, you can obtain the solar azimuth and altitude from this site. Solar altitude or elevation is used to derive solar zenith angle (the sum of solar zenith angle and solar elevation angle is 90 degrees). Solar zenith angle is also sometimes provided by the instrument's telemetry (e.g. EO-1).

Earth-Sun distance (*d*) is entered in Astronomical Units (AU) and is different for each Julian day. This value is known for many days of the year, and can be interpolated for other days from those known values (see Table 1 here). The complete table of distances for all days of the year is available here.

Solar Exoatmospheric Irradiances (E_{SUN_λ}) or solar flux, in (watts / ($m^2 \times \mu m$)), is a function of the wavelength λ at which the measurement is taken. Solar flux values for different Hyperion bands are listed under Table 3 here. For Landsat-7, solar flux values are in Table 9.1 here.

Finally, the cloud detection algorithm is applied to the reflectance values of a few bands of each data set. The algorithm consists of calculating some ratios and checking them against some thresholds for each pixel and determining whether or not that pixel should be classified as cloud or not. This is usually done through an elimination process (determining the pixel is not water, vegetation, fire, etc.). For the remaining of this document, we discuss cloud detection algorithms that are developed for Landsat-7 and Hyperion data.

2. Landsat-7 Cloud Detection Algorithm

The Automatic Cloud Cover Assessment (ACCA) was developed for Landsat-7 data (El-Araby et. al., 2009). This algorithm takes bands 2-6 of Landsat 7 as input. The following preprocessing steps are first applied:

1) Bands 2-6 irradiance values are normalized first as described in section 2.2, Equation 1, and Table.2 of (El. Araby *et al.* 2009).

2) Normalized bands 2-5 irradiance values are converted to reflectance values as described in Section 1, Equation 1, as well as Section 2.2 of the paper by El. Araby *et al 2009.*

3) Band 6 normalized irradiance values are converted to at-sensor temperature using Equation 3 described in Section 2.2 of the paper by El. Araby *et al.* 2009.

From here on whenever we mention Bands 2-5 values we are referring to the results obtained from step 2 above, and whenever we mention Band 6, results obtained from step 3 are used. The algorithm detects clouds in two passes. During the first pass, clouds are differentiated from non-clouds by applying 8 different filters. This algorithm and the logic used for each filter are demonstrated in Figure 2.1. At the end of this process, each pixel in the image gets one of the four possible labels: non-cloud, ambiguous, warm cloud, or cold cloud. Then, pass one detection ambiguities are resolved during the second pass of the ACCA algorithm. In particular, during the second pass a new thermal threshold is applied. Here are the steps taken in pass two:

1) Band 6 statistics including mean, standard deviation, distribution skewness, and kurtosis are calculated (see El Araby *et al.* 2009, Equation block 4, for details).

2) The smallest thermal value that is larger than the 95% of the numbers in the given set of pixels (the 95th percentile), becomes the new thermal threshold. Image pixels that fall below this thermal threshold and pass the following three conditions are classified as cloud pixels.

   a. Desert index ($\frac{B_4}{B_5}$) is greater than 0.5
   b. Colder cloud population exceeds 0.4 percent of the scene
   c. Mean temperature of the cloud class is less than 300K

3) Each non-cloud pixel in the cloud mask is compared with its neighbors. If 5 out of its 8 neighbors are clouds, the pixel is reclassified as cloud. This is done to fill the cloud holes.

4) Cloud mask results obtained from pass one and pass two are compared. Extreme differences indicate cloud signature corruption. In this case, pass-two results are ignored and pass-one results are used.

Note 1: the ACCA algorithm assumes all bands have the same resolution. This is while band 6 data of Landsat-7, acquired before February 25, 2010, has a resolution of 60 meters per pixel. Therefore, one needs to resample band 6 data to increase its resolution to 30 meters per pixel before being able to run ACCA on it. Interpolation techniques such as bilinear interpolation, nearest neighbors, etc. are typically used for this step. If the Landsat-7 image is dated after Feb. 25, 2010, band 6 has already been resampled to 30m resolution than thus there is no need for resampling.

Note 2: the ACCA algorithm includes step-2 mentioned before regarding the conversion of the irradiance values to reflectance values. In addition it performs normalization for its thermal band 6.

Note 3: In theory, one could apply the same ACCA algorithm on input bands that are obtained at the same wavelengths as Landsat-7 bands 2-6, as listed in Table 2.1, were obtained.

Band Number	Wavelength (micron)	Spatial Resolution (meters)
2	0.52-0.60	30
3	0.63-0.69	30
4	0.77-0.90	30
5	1.55-1.75	30
6 (Thermal IR)	10.40-12.50	60 (30 after Feb. 25, 2010)

5) **Table 2.1- Landsat-7 bands used for cloud detection in ACCA algorithm**

Figures 2.2 (a-f) demonstrate the performance of ACCA on three subsets of Landsat-7 scenes from Boston, Chesapeake Bay area, and Colorado accordingly.

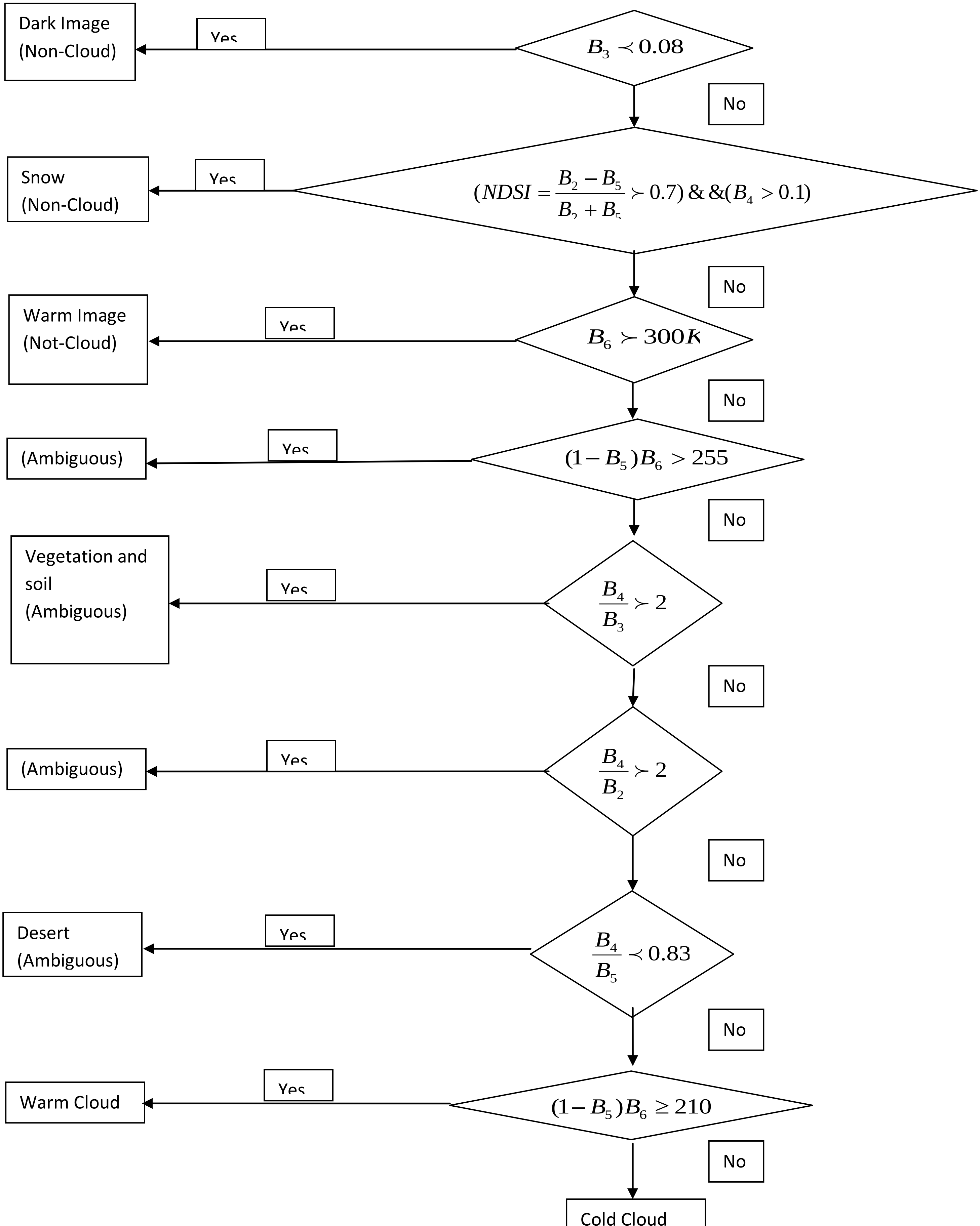


Figure 2.1- Pass one flowchart of the ACCA algorithm that uses eight different filters. At the end of this pass each pixel is labeled as non-cloud, ambiguous, warm cloud, or cold cloud (El-Araby, *et al*. 2009).

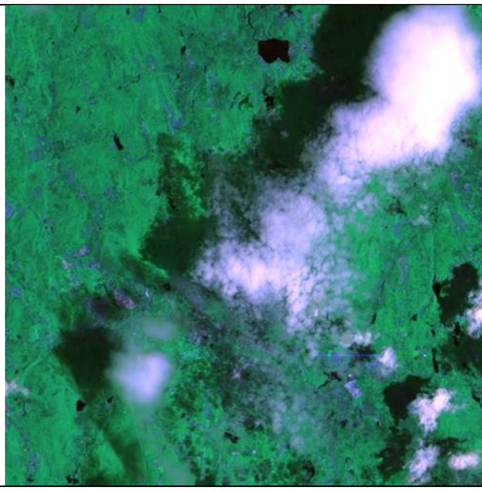
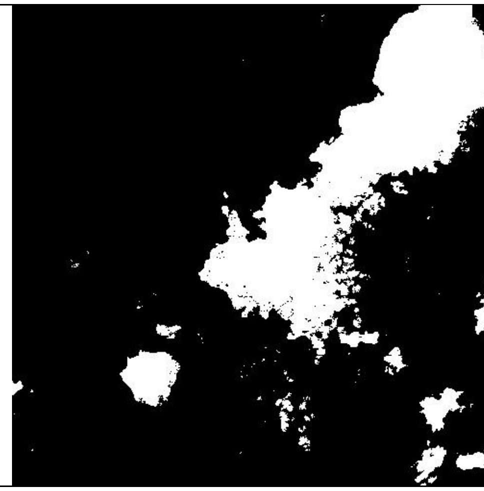
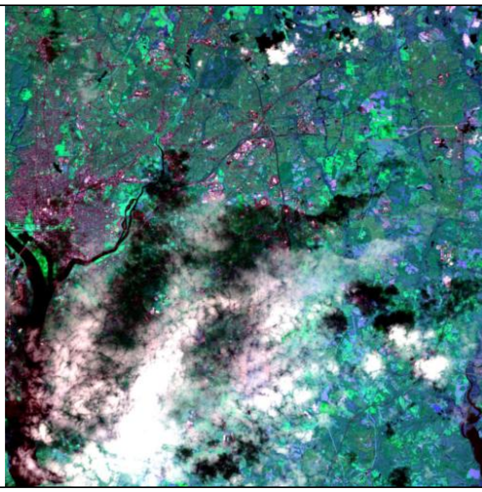
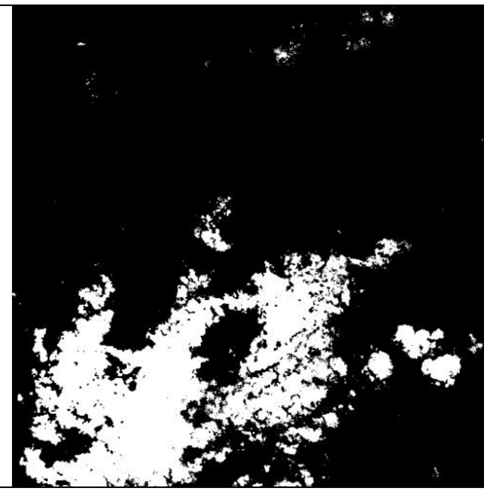
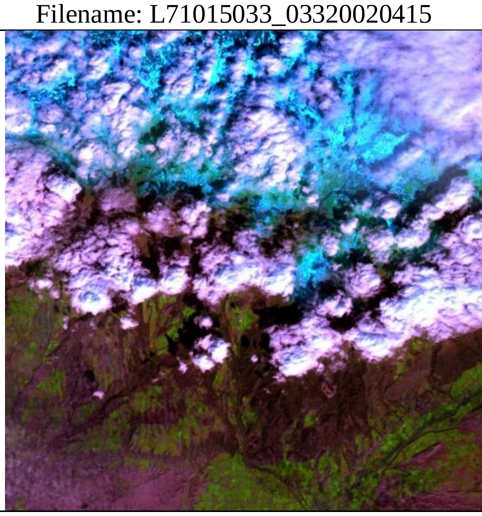
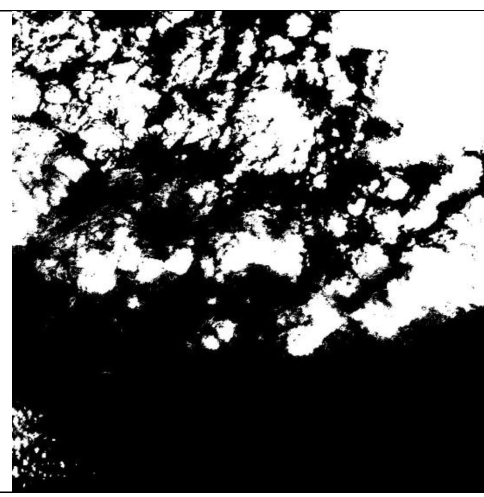
	
a) Landsat Image of the Boston Area, (Original image used in ACCA paper)	b) Cloud Mask- cloud coverage: 18.31%
	
c) Landsat-7 Image subset of the Chesapeake Bay Area Obtained on April 15, 2002. Filename: L71015033_03320020415	d) Cloud Mask- cloud coverage: 18%
	
e) Landsat-7 Image subset of the Colorado Obtained on October 04, 2002. Filename: L71035033_03320021004	f) Cloud Mask- cloud coverage: 33.92%

Figure 2.2: Results of the ACCA algorithm ran on Landsat-7 images a, c, e are shown in images b, d, and e respectively.

3. Hyperion Cloud Detection Algorithm

Earth Observing 1 (EO-1) satellite consists of two instruments: Hyperion and ALI. ALI's spectral coverage is very similar to Landsat-7, except that it misses the corresponding band to the thermal band (band 6) of Landsat-7. Otherwise, one could have used the ACCA algorithm for cloud detection on ALI data by using the bands that correspond to the same Landsat-7 wavelengths that are used in the ACCA, as described above. Hyperion is the hyper-spectral instrument of EO-1 with 242 bands. There is already a cloud detection algorithm running on-board of EO-1 that uses Hyperion data (Griffin *et al.* 2003). The objective of this section is to help readers who are interested in coding the cloud mask algorithm for Hyperion data to understand the algorithm better and to warn them of the issues they should be mindful of when developing their cloud detection programs. This section is based on two published papers on this subject (Griffin *et al.* 2003 and Doggett *et al.* 2006), and the author's communication with a few co-authors of the referenced papers.

Note 1: While Hyperion has 242 bands, the cloud detection algorithm described in (Griffin *et al.* 2003) uses only six of the Hyperion bands, which are bands number 21, 31, 51, 110, 123, and 150 (See Table 3.1).

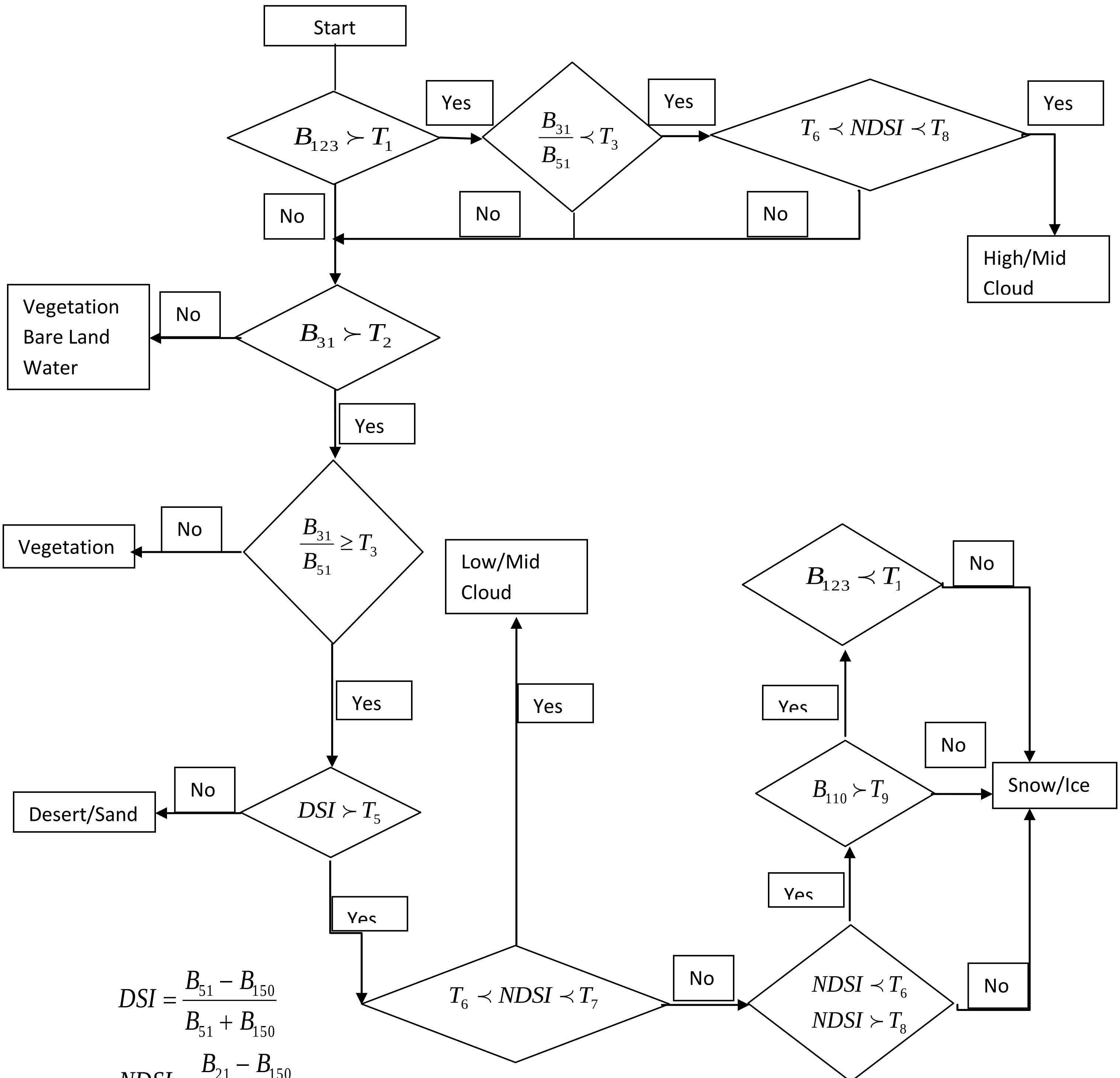
Note 2: The Hyperion algorithm assumes the data values are reflectance values, not radiance values (see Section 1.1, step 2 for details of this conversion).

Note 3: The algorithm, as described in Figure 1 of (Griffin *et al.* 2003) calculates some ratios. It is important to check for their denominators first. If the denominator is zero, that pixel will be labeled unclassified.

Band Number	Central Wavelength Values (micron)	Spatial Resolution (meters)
21	0.56	30
31	0.66	30
51	0.86	30
110	1.25	30
123	1.38	30
150	1.65	30

Table 3.1- Hyperion bands used for cloud detection in Griffin *et al.* algorithm.

Figure 3.1 demonstrates the flowchart of the Hyperion cloud detection algorithm of Griffin *et al.*, 2003. This algorithm uses various threshold comparisons. These threshold values for reflectance values of Hyperion are described in Table 3.2. Experimental results had implied that the algorithm performed better if different thresholds were used depending on whether the algorithm was performed on land data or water data. Recommended thresholds for these two different sets of data are listed in Table 3.2. Figures 2.2 (a-f) demonstrate the performance of ACCA on three subsets of Landsat-7 scenes from Boston, Chesapeake Bay area, and Colorado accordingly. Figures 3.2 demonstrates the performance of this algorithm on a subset of a Hyperion scene from the Tibet area obtained in March 2004.



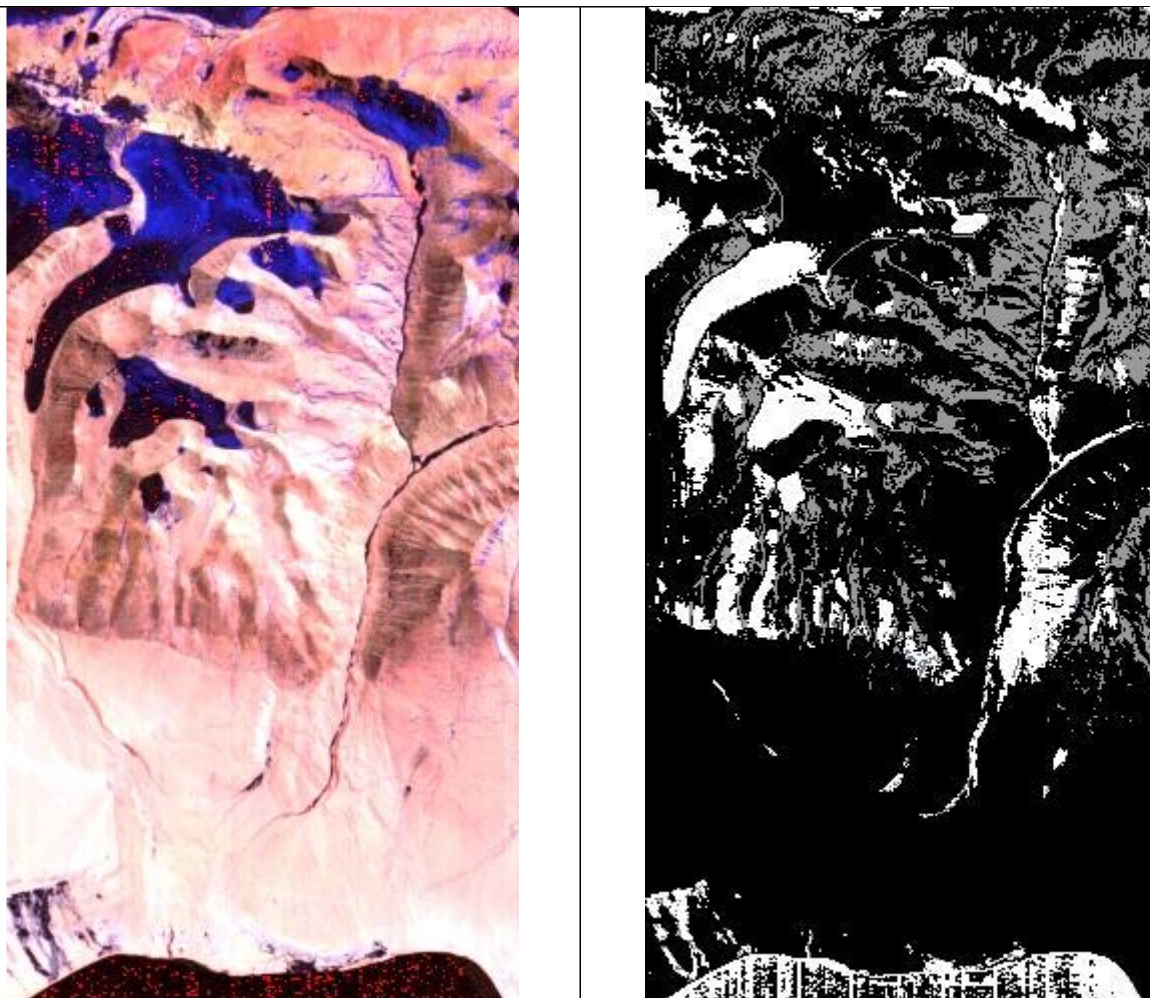
$$DSI = \frac{B_{51} - B_{150}}{B_{51} + B_{150}}$$

$$NDSI = \frac{B_{21} - B_{150}}{B_{21} + B_{150}}$$

Figure 3.1- Flowchart of the Hyperion cloud detection algorithm that uses different filters. At the end each pixel is labeled as unclassified (non-cloud), low cloud, mid cloud, or high cloud (Griffin et al. 2003).

Thresholds	Over Land	Over Water
T_1 : High cloud reflectance threshold	0.10	0.10
T_2 : Red reflectance threshold	0.15	0.15
T_3 : Vegetation ratio lower threshold	0.70	0.60
T_4 : Vegetation ratio upper threshold	1.00	1.00
T_5 : DSI threshold	0.05	0.01
T_6 : NDSI lower threshold	0.00	-0.2
T_7 : NDSI upper threshold	0.20	0.20
T_8 : NDSI extended threshold	0.60	0.60
T_9 : SWIR reflectance threshold	0.35	0.35

Table 3.2- Threshold values for Hyperion cloud detection algorithm in Figure 3.1



Original Hyperion

Cloud Mask: unclassified/low/mid/high clouds.

Figure 3.2 Results of the Hyperion cloud detection algorithm ran on a subset of Hyperion data obtained from Mapam Lake of Tibet in March 2007.

4. References:

1) M. Griffin, H. Burke, D. Mandl, and Jerry Miller. "Cloud Cover Detection Algorithm for EO-1 Hyperion Imagery" in 2003 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), July 2003, vol. 1, pages: 86-89.

   Clout Detection

2) T. Doggett, R. Greeley, S. Chien, R. Castano, B. Cichy, A. G. Davies, G. Rabideau, R. Sherwood, D. Tran, V. Baker, J. Dohm, and F. Ip, "Autonomous detection of cryospheric change with hyperion on-board Earth Observing-1", Remote Sensing of Environment, Vol. 101, Issue 4, April 2006, pages 447-462.

   Doggett_et_al

3) Correspondences with Thomas Doggett, January-April, 2008.
4) Correspondences with Michael K. Griffin, MIT Lincoln Laboratory, August 2010.
5) EO-1 Website (http://eo1.gsfc.nasa.gov/)
   a. EO-1 Preliminary Technology and Science Validation Report
      i. http://eo1.gsfc.nasa.gov/new/validationReport/index.html#part14
      ii. Cloud detection related publications and presentations listed on above link under Part 7 (Sensor Web/ Test bed Initiatives, Section 5.
6) General Overview of the Hyperion's cloud detection algorithm: http://eo1.gsfc.nasa.gov/new/validationReport/Technology/SensorWebs/Final%20Rpt_Appendix_Cloud%20Cover%20Val.ppt
7) E. El-Araby, T. El-Ghazawi, J. Le Moigne, and R. Irish, "Reconfigurable Processing for Satellite On-Board Automatic Cloud Cover Assessment (ACCA)", Journal of Real Time Image Processing, 2009, Vol. 4, No. 3, pp. 245-259.

   Landsat-CloudDetect ion

8) General Overview of the ACCA algorithm: http://landsathandbook.gsfc.nasa.gov/pdfs/ACCA_slides.pdf
9) Landsat Handbook (http://landsathandbook.gsfc.nasa.gov/)
   a. ACCA Algorithm:
      i. http://landsathandbook.gsfc.nasa.gov/htmls/acca.html
      ii. http://landsathandbook.gsfc.nasa.gov/pdfs/ACCA_slides.pdf
10) Landsat facts: http://geo.arc.nasa.gov/sge/landsat/l7.html
11) Sun or moon altitude/azimuth table for U.S. cities: http://www.usno.navy.mil/USNO/astronomical-applications/data-services/alt-az-us
12) USGS FAQ page on conversion of radiance to reflectance values: http://edcsns17.cr.usgs.gov/eo1/faq.php?id=21